Nonconvex compressive sensing for X-ray CT: an algorithm comparison

Rick Chartrand
Los Alamos National Laboratory
rickc@lanl.gov

Emil Y. Sidky
University of Chicago
sidky@uchicago.edu

Xiaochuan Pan
University of Chicago
xpan@uchicago.edu

***Abstract*—Compressive sensing makes it possible to reconstruct images from severely underdetermined linear systems. For X-ray CT, this can allow high-quality images to be reconstructed from projections along few angles, reducing patient dose, as well as enable other forms of limited-view tomography such as tomosynthesis. Many previous results have shown that using nonconvex optimization can greatly improve the results obtained from compressive sensing, and several efficient algorithms have been developed for this purpose. In this paper, we examine some recent algorithms for CT image reconstruction that solve nonconvex optimization problems, and compare their reconstruction performance and computational efficiency.**

I. Introduction

The chief result of compressive sensing (CS) [1], [2] is that we can reconstruct images from very few measurements by exploiting sparsity. Most CS research applies only to linear measurements, which have to satisfy an *incoherence* property. Consequently, X-ray computed tomography (CT) has been a natural application area for CS [3]–[8]: it is well approximated by a linear model, and the Fourier-slice theorem [9] suggests that radiographic measurements are roughly tantamount to sampling in the Fourier domain.

The most common approach to CS has been the use of convex optimization. In particular, the ℓ^1 norm can be an effective penalty function for promoting sparse solutions to optimization problems, while providing the algorithmic benefits of being a convex function. However, many results [6], [10]–[17] have shown improved results using nonconvex penalty functions instead. For X-ray CT, the potential benefits of using nonconvex optimization include reconstruction using fewer radiographic views, and improved robustness to noise, which would allow the total radiation dose to be decreased.

Image reconstruction using CS necessitates the use of iterative algorithms, which can be computationally expensive. As a result, there has been much research in the development of efficient algorithms for CS. Many algorithms for convex optimization have been extended to the nonconvex case [10], [12], [13], [16]. In this paper, we compare a selection of three such algorithms to a synthetic problem in X-ray CT image reconstruction. We discuss these algorithms and issues regarding their implementation in Sec. II. We compare in Sec. III-A the ability of the algorithms to reconstruct a breast tissue phantom from ideal, noise-free data, as well as their efficiency, to assess their potential abilities in reconstructing images from limited views. In Sec. III-B, we consider the case of noisy data, to test algorithm robustness and efficiency.

II. Three algorithms for exploiting gradient sparsity

We will consider the task of reconstructing a synthetic breast tissue phantom (see Fig. 2). Although this phantom has complex edges, its gradient is still fairly sparse. Our approach will be to exploit this sparsity by means of optimization problems designed to promote a sparse gradient. These will take one of the following two forms:

$$\min_{\mathbf{x}} G(\nabla \mathbf{x}) + \tfrac{\mu}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2; \text{ or} \tag{1}$$

$$\min_{\mathbf{x}} G(\nabla \mathbf{x}) \;\text{ subject to }\; \|A\mathbf{x}-\mathbf{b}\|_2 \le \sigma. \tag{2}$$

Here G will be one of several penalty functions whose minimizers tend to be sparse. The best-known example is the ℓ^1 norm, $G(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$, in which case $G(\nabla\mathbf{x})$ is the total variation (TV) of $\mathbf{x}$. We seek to reconstruct $\mathbf{x} \in \mathbb{R}^N$, which contains the attenuation values at each pixel. Our ∇ is a discrete approximation of the continuous gradient. The radiographic data are stored in $\mathbf{b}$, and A is the 2-D fanbeam transform. The parameters μ or σ balance the competing effects of regularization of the image $\mathbf{x}$, and consistency of the measurements $A\mathbf{x}$ with the data $\mathbf{b}$. Although problems (1) and (2) are equivalent (for corresponding choices of μ and σ), in practice it is easier to select σ, as this corresponds to the combined effects of noise and model inaccuracy, while μ does not have a direct, physical interpretation.

A. Iteratively reweighted least squares

The origins of iteratively reweighted least squares (IRLS) methods go back to at least the 1960s [18]. It was first applied to (1) in the case of TV by Vogel and Oman [19]. We will apply the method to the case of the ℓ^p norm: $G(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum_i |x_i|^p$, where $0 < p < 1$ [20]. The idea is to approximate the penalty function by a weighted quadratic function, using the previous iterate to compute the weights:

$$\|\nabla\mathbf{x}\|_p^p \approx \sum_{i=1}^N w_i^n |(\nabla\mathbf{x})_i|^2, \tag{3}$$

where at the n^{th} iteration, $w_i^n = \left(\sqrt{|(\nabla\mathbf{x}^n)_i|^2+\epsilon}\right)^{p-2}$. We introduce $\epsilon > 0$ to avoid division by zero. Substituting (3) into (1), we obtain a quadratic optimization problem, which

results in a linear system. Manipulation of this system in the manner done in [19] produces an iteration that has a structural resemblance to a quasi-Newton method:

$$\begin{aligned}\big(\nabla^T \operatorname{diag}(\mathbf{w}^n)\nabla + \mu A^T A\big)\big(\mathbf{x}^{n+1} - \mathbf{x}^n\big) \\ = \mu A^T \mathbf{b} - \big(\nabla^T \operatorname{diag}(\mathbf{w}^n)\nabla + \mu A^T A\big)\mathbf{x}^n. \quad (4)\end{aligned}$$

Using a larger value of ϵ than needed to avoid division by zero can make the linear system in (4) better conditioned, which makes the iteration faster. If ϵ is too large, this effect can come at the expense of accuracy of the solution. A continuation strategy, where ϵ is initially large and then decreased as the iteration proceeds, can be used to balance these issues. It also turns out that such a strategy can help the iteration avoid local minima when $p < 1$ [12]. Heuristically, the local minima are smoothed away when ϵ is large, and then when it is decreased and local minima re-appear, the iteration can already be in the right basin of convergence. Convergence has only been established in the absence of our analysis operator ∇, and then only to a local minimizer when $p < 1$ [21].

B. Alternating direction method of multipliers

The alternating direction, method of multipliers (ADMM) algorithm dates to the 1970s [22], [23]. It was re-discovered in the context of image processing in 2008 [24], [25], and generalized to the nonconvex case in [16]. Applied to (1), we introduce a variable $\mathbf{w}$ to take the place of $\nabla\mathbf{x}$, and a term penalizing their difference. This decomposes the problem into two, much simpler subproblems. In order to enforce convergence of $\mathbf{w}$ to $\nabla\mathbf{x}$, we employ the method of multipliers [26], [27], and introduce a dual variable Λ, to which the residual is added at each iteration. In the noise-free case, we can do the same to enforce $A\mathbf{x} = \mathbf{b}$ at convergence. The optimization problem now takes the following form:

$$\min_{\mathbf{w},\mathbf{x}} G(\mathbf{w}) + \tfrac{1}{2\lambda}\|\mathbf{w} - \nabla\mathbf{x} - \Lambda\| + \tfrac{\mu}{2}\|A\mathbf{x} - \mathbf{b} - \Lambda_2\|_2^2, \quad (5)$$

where we omit Λ_2 in the noisy case. The iteration becomes:

$$\begin{aligned}\text{Solve } (\tfrac{1}{\lambda}\nabla^T\nabla + \mu A^T A)\mathbf{x}^{n+1} \quad (6)\\ = \tfrac{1}{\lambda}\nabla^T(\mathbf{w}^n - \Lambda_2^n) + \mu A^T(\mathbf{b} + \Lambda_1^n), \quad (7)\\ \mathbf{w}^{n+1} = S^\lambda(\nabla\mathbf{x}^{n+1} + \Lambda_2^n), \quad (8)\\ \Lambda^{n+1} = \Lambda^n + \gamma(\mathbf{b} - A\mathbf{x}^{n+1}), \quad (9)\\ \Lambda_2^{n+1} = \Lambda_2^n + \gamma(\nabla\mathbf{x}^{n+1} - \mathbf{w}^{n+1}). \quad (10)\end{aligned}$$

Here γ is a parameter that can be used to accelerate the method of multipliers by using $\gamma > 1$. In the convex case, it is known that γ can be as large as $(\sqrt{5}+1)/2 \approx 1.618$ [28]. In this paper we use $\gamma = 1.6$, giving us guaranteed convergence when $p = 1$, with only empirically observed convergence when $p < 1$.

In (8), S^λ is the *proximal mapping* of G, defined as follows:

$$S^\lambda(\mathbf{x}) = \arg\min_{\mathbf{w}} G(\mathbf{w}) + \tfrac{1}{2\lambda}\|\mathbf{w} - \mathbf{x}\|_2^2. \quad (11)$$

When G is the ℓ^1 norm, S^λ is known as soft thresholding. More generally, we will consider the following mappings for

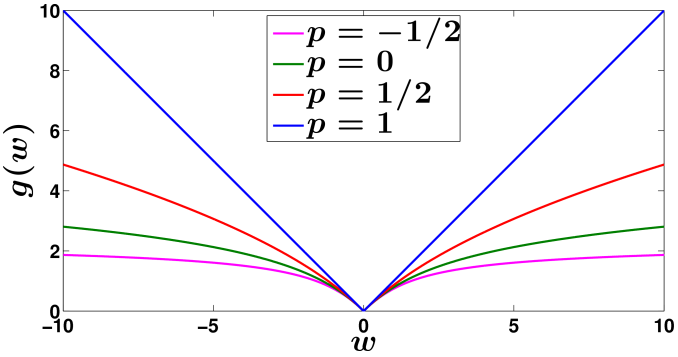


Fig. 1: Plots of $g = g_p$, where $G_p(\mathbf{w}) = \sum_i g_p(|w_i|)$ is the penalty having S_p in (12) for a proximal mapping. $g_p(|w|)$ grows like $|w|^p/p + C$, but has different behavior from a p^{th} power function near $w = 0$.

$p \le 1$, defined componentwise:

$$S_p^\lambda(\mathbf{x})_i = \max\{0, |x_i| - \lambda^{2-p}|x_i|^{p-1}\}\operatorname{sign}(x_i). \quad (12)$$

Then soft thresholding is the case of $p = 1$. For $p < 1$, we use (11) to *define* $G = G_p$, where we use $S = S_p$. The details of how G_p can be so constructed are in [29] (see also [17]). The resulting penalty function G_p can be regarded as a modification of the ℓ^p norm (to the p^{th} power), with the modification being for the purpose of having a simple proximal mapping for use in (8) (a property the ℓ^p norm itself does not have). We do not have a formula for G_p for general p, but we do not need one for our algorithm. We can compute G_p numerically, with plots of the 1-D case in Fig. 1.

C. Chambolle-Pock

In [30], Chambolle and Pock develop an approach that can applied fairly broadly to convex minimization problems, and can be seen as generalizations of earlier work [31], [32]. Underlying the approach is the incorporation of duality. Consider (1) with G being the ℓ^1 norm. Duality gives us the following "saddle-point" expression:

$$\min_{\mathbf{x}} \|\nabla\mathbf{x}\|_1 + \tfrac{\mu}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 = \min_{\mathbf{x}} \max_{\|\mathbf{y}\|_\infty \le 1} \langle\nabla\mathbf{x}, \mathbf{y}\rangle + \tfrac{\mu}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2. \quad (13)$$

Since $\langle\nabla\mathbf{x}, \mathbf{y}\rangle = \langle\mathbf{x}, \nabla^T\mathbf{y}\rangle$, each $\mathbf{x}$ and $\mathbf{y}$ can be considered without the presence of an analysis operator. Then *proximal gradient* methods can be applied, leading to a *primal-dual* algorithm.

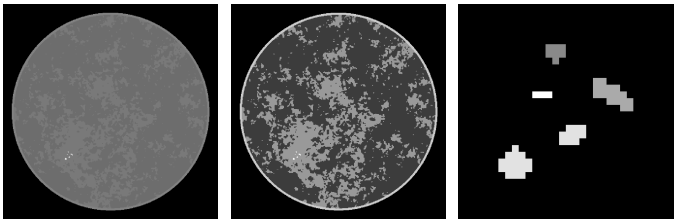
The Chambolle-Pock framework can only be directly applied to convex problems. We consider the case of G being the ℓ^p norm (to the p^{th} power) by iteratively reweighting the ℓ^1 norm:

$$\begin{aligned}\min_{\mathbf{x}} \sum_i w_i|(\nabla\mathbf{x})_i| + \tfrac{\mu}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 \\ = \min_{\mathbf{x}} \max_{\forall i(|y_i| \le w_i)} \langle\nabla\mathbf{x}, \mathbf{y}\rangle + \tfrac{\mu}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2,\end{aligned}$$

with $w_i = |(\nabla\mathbf{x}^n)_i|^{p-1}$, computed from the previous iterate.

We obtain better results by considering the constrained formulation (1). This can be handled within the Chambolle-Pock framework by using an indicator function:

$$\min_{\mathbf{x}} \|\nabla\mathbf{x}\|_p^p + \delta_{\{\mathbf{x}:\|A\mathbf{x}-\mathbf{b}\|_2 \le \sigma\}}(\mathbf{x}), \quad (14)$$



(a) 512×512 (b) enhanced contrast (c) $16\times$ zoom

Fig. 2: The breast tissue phantom used in our experiments. In (b) the display window is narrowed to show the soft tissues, while (c) zooms in on the microcalcifications.

where the indicator function δ gives $\mathbf{x}$ the value 0 if $\mathbf{x}$ is feasible, otherwise $+\infty$. Referring to [30, p. 121], the entire functional in (14) becomes $F(K\mathbf{x})$ (with K being a stack of A and ∇), while their G is identically zero.

Combining these ingredients, we arrive at the following iteration, first appearing in [33]. Since we solve a constrained problem for minimizing a "total p-variation," we refer to the algorithm as CTpV.

$$\mathbf{y}^{n+1} = \operatorname{prox}_{\alpha\sigma} \|\cdot\|_2(\mathbf{y}^n + \alpha(A\mathbf{x}^n - \mathbf{b})), \tag{15}$$

$$\mathbf{w}^{n+1} = \left(\sqrt{|\nabla \mathbf{x}^n|^2 + \epsilon^2}/\epsilon\right)^{p-1}, \tag{16}$$

$$z_i^{n+1} = P_{B^\infty(\lambda w_i^{n+1}/\nu_i)}(z_i^n + \alpha\nu(\nabla \mathbf{x}^n)_i), \tag{17}$$

$$\mathbf{x}^{n+1/2} = \mathbf{x}^{n-1/2} - \beta(A^T\mathbf{y}^{n+1} + \nu\nabla^T\mathbf{z}^{n+1}), \tag{18}$$

$$\mathbf{x}^{n+1} = \mathbf{x}^{n+1/2} + \theta(\mathbf{x}^{n+1/2} - \mathbf{x}^{n-1/2}); \tag{19}$$

where $\operatorname{prox}_\mu \|\cdot\|_2(\mathbf{y}) = \max\{\|\mathbf{y}\|_2 - \mu, 0\}\mathbf{y}/\|\mathbf{y}\|_2$, the proximal mapping of the ℓ^2 norm. The second step incorporates an $\epsilon > 0$ both to avoid division by zero and provide some regularity. The third step uses orthogonal projection onto the ℓ^∞ ball of the specified radius. The iteration is provably convergent only when $p = 1$, or with fixed weights.

III. Numerical Experiments

We consider the synthetic breast-tissue phantom in Fig. 2. It is an instance of a probabilistic model from [34]. Present are fat, fibroglandular tissue, skin, and microcalcifications. It is discretized on a 512×512 grid measuring 18 cm on a side. We call the vectorized phantom $\mathbf{x}^*$.

Our X-ray transform A implements a 2-D fanbeam transform. The source is 36 cm from the center of the object, and 72 cm from the center of the detector. The detector is linear with 1024 bins and total length 36 cm. The transform is right-multiplied by a diagonal matrix of 0s and 1s, with 0s corresponding to pixels whose centers are outside the circle inscribed in the grid, as only the pixels in this circle are allowed to vary. (Note that this circle is larger than the support of the object.) The resulting A is constructed in Matlab as a sparse matrix, using code provided to the author by Jakob S. Jørgensen as a supplement to AIR Tools [35].

Our gradient operator ∇ is a discrete, finite-difference approximation. We use simple forward differencing in each dimension. It is implemented using a pair of sparse matrices, one for the derivative in each dimension. Each is masked in the same manner as A, to ensure only the pixels in the inscribed circle can vary. This also has the effect of making boundary conditions for the finite differencing irrelevant.

A. Noise-free case

First we consider the ideal-data case, where $\mathbf{b} = A\mathbf{x}^*$. This is to investigate the algorithms' capabilities with regard to limited views, independently from other confounding issues. We present results for A corresponding to 30, 35, 40, and 45 views, all equally spaced in $[0, 2\pi]$. The choice was informed by the results: 45 is sufficiently many views to make a high-quality reconstruction possible using all considered values of p, while 25 views was not found to be sufficient in any case.

In this ideal case, our objective it to obtain an essentially perfect reconstruction. We regard a reconstruction SNR of 50 dB as sufficiently high quality to meet this objective. We compare the three algorithms according to whether they attain our reconstruction threshold, and if so, in how much time. The computations are done in Matlab on a common workstation. We consider $p \in \{1/4, 1/2, 3/4, 1\}$. Each algorithm has two parameters to choose: μ and ϵ for IRLS, μ and λ for ADMM, and λ and ϵ for CTpV. (The other CTpV parameters α, β, θ, ν, and σ all can be chosen *a priori* as described in [33].) The parameters are chosen to optimize the reconstruction time. For ADMM and CTpV, the parameters are chosen with a two-dimensional, logarithmic grid search, with grid spacing $10^{0.1}$. For IRLS in the noise-free case, this is not possible, as both parameters must be adapted with a continuation strategy as the iteration proceeds. Reasonable parameter sequences were sought, generally with factor of 10 increases in μ and decreases in ϵ, but it was infeasible to consider all possibilities, so no claim of optimality can be made.

The results are in Table I. We see that the results for CTpV are uniformly better than those for ADMM and IRLS. Reconstruction was successful for CTpV except in cases where it could not have possibly succeeded because not enough measurements were made for the phantom to be the minimizer of the optimization problem (these cases being $p = 1$ with 40 or fewer views). Results across all algorithms were generally better for $p = 1/2$ or $p = 3/4$. IRLS both failed in the most cases, and took longest where it succeeded. ADMM was mostly as reliable as CTpV, but much slower, something exacerbated by its tendency to oscillate. Note that both IRLS and ADMM require a linear system to be solved at each iteration, while CTpV does not, this perhaps being the main reason for the much faster times. (In smaller problems, ADMM can benefit from having the system matrix in (7) being fixed, allowing a single Cholesky factorization, with only a backsubstitution necessary at each iteration. In this case, A^TA, being dense, is too large to construct, let alone factor.)

B. Noisy case

For the noisy case, A corresponds to 50 views. We apply synthetic Poisson noise to simulated exponential attenuation data. That is, we compute $\exp(-\alpha A\mathbf{x}^*)$, apply Poisson noise to obtain detector data $\mathbf{d}$, then compute $\mathbf{b} = -\log(\mathbf{d})/\alpha$. The scaling constant α is computed so that the total incident photon

	IRLS				ADMM				CTpV			
views	1/4	1/2	3/4	1	1/4	1/2	3/4	1	1/4	1/2	3/4	1
45	fail	471	2066	fail	661	373	292	229485	136	**116**	116	6212
40	fail	907	2992	break	851	535	431	break	170	150	**139**	break
35	fail	1476	6593	break	842	548	722	break	198	171	**153**	break
30	fail	fail	fail	break	1555	1272	fail	break	320	296	**292**	break

TABLE I: Time in seconds to reconstruct to an SNR of 50 dB, for three different algorithms, four values of p, and four numbers of views. 'break' indicates the solution proved that the phantom is not the global minimizer, by having a lower objective value and being feasible. 'fail' indicates the algorithm failed to reconstruct the phantom sufficiently well, without the solution having a lower objective value. The CTpV algorithm was fastest for every number of views (table entries indicated in bold), as well as fastest for any particular value of p, while also only failing in cases where the phantom was demonstrably not the minimizer of the optimization problem.

	time (s)			SNR (dB)		
p	IRLS	ADMM	CTpV	IRLS	ADMM	CTpV
1/4	479	1404	149	26.2	26.0	**27.1**
1/2	402	1407	**83**	26.7	25.9	27.1
3/4	358	1438	156	26.8	25.8	26.8
1	384	1336	131	26.0	25.4	26.0

TABLE II: Reconstruction times and SNRs for the noisy-data case. CTpV results have best SNR, and by far the fastest times. Quality is mostly better with smaller p, while $p = 1/2$ or 3/4 gives the fastest times.

flux is 264,000 photons per view. Note that this is four times as many photons per view as in [33], which used 200 views, keeping the total number of photons the same, and equivalent to two-view, full-field digital mammography [36].

The experiments were conducted similarly to the noise-free case, with a few exceptions. In the absence of a task for which the images will be used, parameters were chosen to produce the highest SNR over the whole image. Due to the softer constraint, continuation on μ was not found to be necessary for IRLS, so the same logarithmic grid search as in Sec. III-A was done for μ. For CTpV, the inequality constraint parameter σ now also must be chosen. Since our focus is on algorithm capabilities, the initial value for σ was the oracle value $\|A\mathbf{x}^* - \mathbf{b}\|_2$. Then the 2-D parameter search for λ and ϵ was conducted as in the noise-free case. Finally, with λ and ϵ fixed, σ was varied in 5% increments, though in all but the $p = 1$ case the improvement was negligible.

Results are in Table II and Fig. 3. While the SNRs do not vary much, CTpV does give the best values. They also give the best subjective appearance: ADMM results appear noisier, while IRLS is blockier with small p. Small artifacts appear present for $p = 1$ with IRLS and CTpV. In Fig. 4 we see the zoom on the microcalcifications, while noting that the parameters were not chosen specifically for this region of interest. Results are mostly better with small p, and with ADMM not doing as well as the other two algorithms. CTpV is by far the fastest. This time, IRLS is second fastest, having less trouble with the softer constraint.

For comparison, in Fig. 5 we see the result of using the classical Kaczmarz algorithm [37], using AIR Tools [35]. The result is noisy, with a much lower SNR of 16.0 dB. The reconstruction took 932 seconds, better than ADMM but much slower than CTpV or IRLS.

IV. Conclusion

Of the three algorithms considered, CTpV produced by far the fastest convergence, while also giving somewhat better robustness to noise. The results also show that the use of nonconvex penalty functions gives better results, though our results do not point to an optimal value of p.

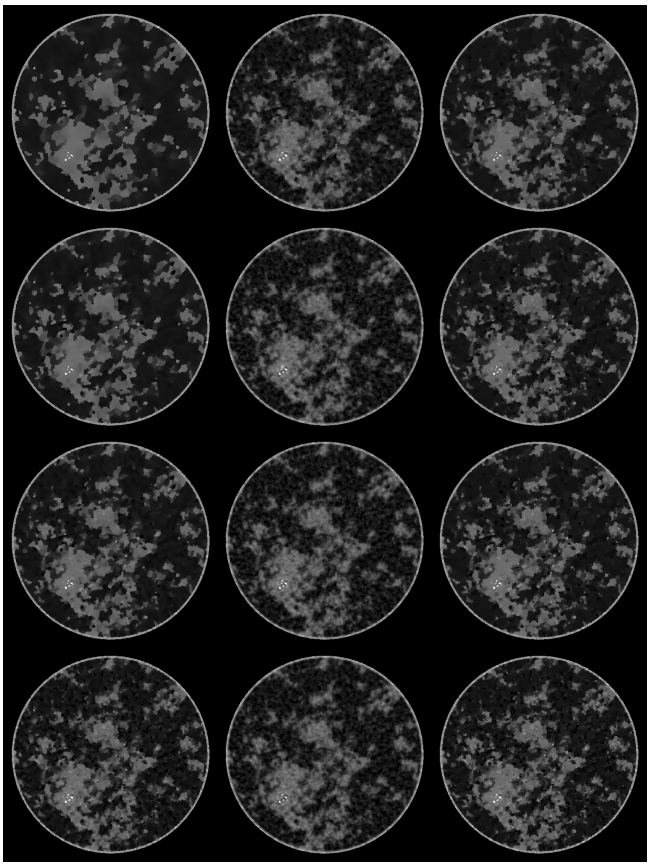


Fig. 3: Reconstructed images, with a uniform display window. The columns are the results using IRLS, ADMM, and CTpV, with the rows being p values of 1/4, 1/2, 3/4, and 1. ADMM results are noisiest, with IRLS blockier for small p. CTpV gives best balance of noise reduction without blockiness. (Best viewed electronically.)

Acknowledgments

The work of RC was supported by the U.S. Department of Energy through the LANL/LDRD Program, and by the University of California Laboratory Fees Research Program. The work of EYS and XP was supported in part by NIH R01 grants CA158446, CA120540, and EB000225. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[3] M. H. Li, H. Q. Yang, and H. Kudo, "An accurate iterative reconstruction algorithm for sparse objects: Application to 3D blood vessel reconstruction from a limited number of projections." *Phys. Med. Biol.*, vol. 47, pp. 2599–2609, 2002.

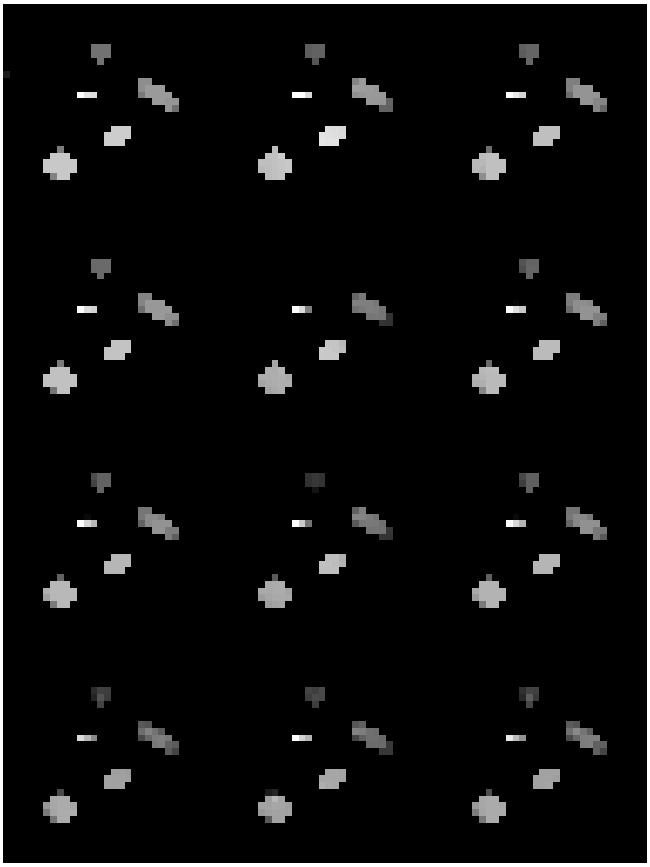


Fig. 4: Same as in Fig. 3, but zoomed in on microcalcifications. ADMM and $p = 1$ have trouble recovering details.

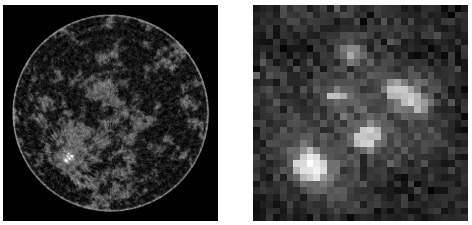


Fig. 5: Results using the Kaczmarz algorithm for comparison. The results are very noisy.

[4] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-ray Sci. Tech.*, vol. 14, pp. 119–139, 2006.

[5] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," *IEEE Trans. Nuc. Sci.*, vol. 34, pp. 4476–4483, 2007.

[6] E. Y. Sidky, R. Chartrand, and X. Pan, "Image reconstruction from few views by non-convex optimization," in *IEEE Medical Imaging Conference Record*, 2007.

[7] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, pp. 660–663, 2008.

[8] J. S. Jørgensen, E. Y. Sidky, and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT," *IEEE Trans. Med. Imaging*, vol. 32, pp. 460–473, 2013.

[9] R. N. Bracewell, "Strip integration in radio astronomy," *Aust. J. Phys.*, vol. 9, pp. 198–217, 1956.

[10] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[11] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[12] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[13] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

[14] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. Imaging*, vol. 28, pp. 106–121, 2009.

[15] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. Signal Process.*, vol. 57, pp. 289–301, 2009.

[16] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging*, 2009.

[17] ——, "Generalized shrinkage and penalty functions," in *IEEE Global Conference on Signal and Information Processing*, 2013.

[18] C. L. Lawson, "Contributions to the theory of linear least maximum approximation," Ph.D. dissertation, University of California, Los Angeles, 1961.

[19] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

[20] R. Chartrand, "Nonconvex regularization for shape preservation," in *IEEE International Conference on Image Processing*, 2007.

[21] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, pp. 1–38, 2010.

[22] R. Glowinski and A. Marrocco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de Dirichlet non lineares," *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, 1975.

[23] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comp. Math. Appl.*, 1976.

[24] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imaging Sci.*, vol. 1, pp. 248–272, 2008.

[25] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, 2009.

[26] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, pp. 303–320, 1969.

[27] M. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed. Academic Press, New York, 1969, pp. 283–298.

[28] *Augmented Lagrangian methods: applications to the numerical solutions of boundary-value problems*. Amsterdam: Elsevier Science Publishers B.V., 1983.

[29] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012.

[30] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, pp. 120–145, 2011.

[31] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Nonlinear Programming*. Stanford: Stanford University Press, 1958.

[32] M. Zhu and T. F. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," UCLA CAM, Tech. Rep. 08-34, 2008.

[33] E. Y. Sidky, R. Chartrand, J. M. Boone, and X. Pan, "Constrained TpV minimization for enhanced exploitation of gradient sparsity: application to CT image reconstruction," 2013, submitted.

[34] I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," *Med. Phys.*, vol. 37, pp. 1591–1600, 2010.

[35] P. C. Hansen and M. Saxild-Hansen, "AIR Tools - a MATLAB package of algebraic iterative reconstruction methods," *J. Comp. Appl. Math.*, vol. 236, pp. 2167–2178, 2012.

[36] J. M. Boone, A. L. C. Kwan, J. A. Seibert, N. Shah, K. K. Lindfors, and T. R. Nelson, "Tehcnique factors and their relationship to radiation dose in pendant geometry breast CT," *Med. Phys.*, vol. 32, pp. 3767–3776, 2005.

[37] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bull. Int. Acad. Pol. Sci. Lett.*, vol. 35, pp. 355–357, 1937.